

BEST AVAILABLE COPY

Source 1..2556
/organism="unknown"
BASE COUNT 797 a 614 c 514 g 631 t
ORIGIN

Query Match 99.1% Score 2534; DB 6; Length 2556;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2556; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 1 GGGAGCCCTGTGTACCAAAAGTGGCTCCGGTCAAGCCGCCCTCTCTAAAGTCAGAGCGGCCCT 60
DB 1 GGGAGCCCTGTGTACCAAAAGTGGCTCCGGTCAAGCCGCCCTCTCTAAAGTCAGAGCGGCCCT 60
QY 61 AGGCTGCTGCTCTCTAGATAGTGGCGGTGAAGCCGCCCAACACACAGACATCACTTT 120
DB 61 AGGCTGCTGCTCTCTAGATAGTGGCGGTGAAGCCGCCCAACACACAGACATCACTTT 120
QY 121 CCGTCAATTTTGAAGCTTCTCCAGAAAGCGTTTGAATTAAGTAAACAGTGGCGCTTG 180
DB 121 CCGTCAATTTTGAAGCTTCTCCAGAAAGCGTTTGAATTAAGTAAACAGTGGCGCTTG 180
QY 181 ATGTGTATCTCCTCAGCAAACTGTAAACAAGCCGAGACACAGATTAACATTAACCTCA 240
DB 181 ATGTGTATCTCCTCAGCAAACTGTAAACAAGCCGAGACACAGATTAACATTAACCTCA 240
QY 241 AGCCGACAGTACAGCGCAATCTCTCAAAACAGTCAAAATCTGTACAGTGGCGAATAGC 300
DB 241 AGCCGACAGTACAGCGCAATCTCTCAAAACAGTCAAAATCTGTACAGTGGCGAATAGC 300
QY 301 TCACATTAATCAAGAAAGTGGCGAGTACACCTTTAAAAATGGGCAAAATAGAACT 360
DB 301 TCACATTAATCAAGAAAGTGGCGAGTACACCTTTAAAAATGGGCAAAATAGAACT 360
QY 361 ACTGTGTAAACACTGTCTCCGAAGCCCTCTCACTACATCTGTGGCTGTGCCAACAGT 420
DB 361 ACTGTGTAAACACTGTCTCCGAAGCCCTCTCACTACATCTGTGGCTGTGCCAACAGT 420
QY 421 RGTGTACAGTTACCTCCTGGAAGCCATGTGAATCTGTAACTACCTGTGAAGCTTAAAG 480
DB 421 -GTGTACAGTTACCTCCTGGAAGCCATGTGAATCTGTAACTACCTGTGAAGCTTAAAG 479
QY 481 TTTGGGAGCATCACTCCTTCAAAATGAGCCCAATCTTAAAGCAGAGAACTCAGAGC 540
DB 481 TTTGGGAGCATCACTCCTTCAAAATGAGCCCAATCTTAAAGCAGAGAACTCAGAGC 539
QY 541 TGTTCAGATTAATCTTCTCCGACATGTGAAGAAATGTGAAGAAATGCAAGAACTTCT 600
DB 541 TGTTCAGATTAATCTTCTCCGACATGTGAAGAAATGTGAAGAAATGCAAGAACTTCT 599
QY 601 TGAATGTTTAAATAAATACATGTAGTGAATGATGATGATGATGATGATGATGATGAT 660
DB 601 TGAATGTTTAAATAAATACATGTAGTGAATGATGATGATGATGATGATGATGATGAT 659
QY 661 GAAGAAGCTGTGGAACAACCTTTGATGCAAAAATGGAAGCAGAGAAATTTACTAGAA 720
DB 661 GAAGAAGCTGTGGAACAACCTTTGATGCAAAAATGGAAGCAGAGAAATTTACTAGAA 719
QY 721 ACTGTATGTTGAACCTCAAGCTTCACTCAGCCCTCACTGCTGCTTCTTTTGAAGAAAG 780
DB 721 ACTGTATGTTGAACCTCAAGCTTCACTCAGCCCTCACTGCTGCTTCTTTTGAAGAAAG 779
QY 781 CGTGTGTGCTTGAAGCAAACTTCTGCTTCACTCAGAGCTTCACTCAGAGATGTGTCA 840
DB 781 CGTGTGTGCTTGAAGCAAACTTCTGCTTCACTCAGAGCTTCACTCAGAGATGTGTCA 839
QY 841 GCAGACTTCTAGTACATGTGTACATGTGTACATGTGTACATGTGTACATGTGTACAT 900
DB 841 GCAGACTTCTAGTACATGTGTACATGTGTACATGTGTACATGTGTACATGTGTACAT 899
QY 901 GGTGACAACTACAGTGTCTCAAGCCAGTCTGAAAGATCAATTAATTTCTGTGAGCAAC 960
DB 901 GGTGACAACTACAGTGTCTCAAGCCAGTCTGAAAGATCAATTAATTTCTGTGAGCAAC 959

QY 961 AGCACCAGAACTGTGTACAGTGAACCTTTGAACCCACTTGTGTGTGAGTGGAGCAAA 1020
DB 961 AGCACCAGAACTGTGTACAGTGAACCTTTGAACCCACTTGTGTGTGAGTGGAGCAAA 1019
QY 1021 AGCTGAGTTGTGACACTTCTCTGTGGGCCCAACTGTGTGACAGAGGAGAAACACAGC 1080
DB 1021 AGCTGAGTTGTGACACTTCTCTGTGGGCCCAACTGTGTGACAGAGGAGAAACACAGC 1079
QY 1081 TGAACATGTTGTGCTTCACTTCAAAACCACTTGTGATGTGTGTGGAACAGATGAC 1140
DB 1081 TGAACATGTTGTGCTTCACTTCAAAACCACTTGTGATGTGTGTGGAACAGATGAC 1139
QY 1141 CAGGTGTCTACATGCAACCTGAAAAGCACTTGTGTGGAACAGATGACATGTCCT 1200
DB 1141 CAGGTGTCTACATGCAACCTGAAAAGCACTTGTGTGGAACAGATGACATGTCCT 1199
QY 1201 TCCAGAGTACATTTTGGAGAAACCTTCAAGTGTGAGTATTTCTTCCATCTGTGAACC 1260
DB 1201 TCCAGAGTACATTTTGGAGAAACCTTCAAGTGTGAGTATTTCTTCCATCTGTGAACC 1259
QY 1261 TGTGTGTCTCTGTGTGAGACATCTGCAAGCTTGTATTTGGAGTCCAGTTCAAAT 1320
DB 1261 TGTGTGTCTCTGTGTGAGACATCTGCAAGCTTGTATTTGGAGTCCAGTTCAAAT 1319
QY 1321 CAAACTTGTCCAGCGGGCCCTGTCTTCAACAACAGCTGGGATTCACAAGGCACTTC 1380
DB 1321 CAAACTTGTCCAGCGGGCCCTGTCTTCAACAACAGCTGGGATTCACAAGGCACTTC 1379
QY 1381 AAGCAGCAACATTTCTCATTTGTTCAGTGTGTGACAGCTTCAAGAGGCAATGAAA 1440
DB 1381 AAGCAGCAACATTTCTCATTTGTTCAGTGTGTGACAGCTTCAAGAGGCAATGAAA 1439
QY 1441 ACAAGTACACCAATTTCTCATTTGTTCAGTGTGTGACAGCTTCAAGAGGCAATGAAA 1500
DB 1441 ACAAGTACACCAATTTCTCATTTGTTCAGTGTGTGACAGCTTCAAGAGGCAATGAAA 1499
QY 1501 GATGCAAGTGAACACATTAATCTGATGATGATGATGATGATGATGATGATGATGAT 1560
DB 1501 GATGCAAGTGAACACATTAATCTGATGATGATGATGATGATGATGATGATGATGAT 1559
QY 1561 AATTTCTGCTGCTGGAATTAATCTGCACTTCAACATCTCTCTCTGAGAAAAA 1620
DB 1561 AATTTCTGCTGCTGGAATTAATCTGCACTTCAACATCTCTCTCTGAGAAAAA 1618
QY 1621 GAATTAAGAGATGTATACATCTGCTCCGAGATGAGATGATGATGATGATGATGAT 1680
DB 1621 GAATTAAGAGATGTATACATCTGCTCCGAGATGAGATGATGATGATGATGATGAT 1678
QY 1681 CTATGCGAGGGGTCAACCTTAATGAAGAAATGCTGATCTTACGACAACTCTGAAT 1740
DB 1681 CTATGCGAGGGGTCAACCTTAATGAAGAAATGCTGATCTTACGACAACTCTGAAT 1738
QY 1741 TGGTGTGCACTCACTTCACTGATGATGAAGTGAAGCAATTTCTTTTATGAGAGCTTC 1800
DB 1741 TGGTGTGCACTCACTTCACTGATGATGAAGTGAAGCAATTTCTTTTATGAGAGCTTC 1798
QY 1801 AAAAGAGATCTTAGACATTTGTAAGAAACATGACATTAACAACTCTGATGCTG 1860
DB 1801 AAAAGAGATCTTAGACATTTGTAAGAAACATGACATTAACAACTCTGATGCTG 1858
QY 1861 TGAATCTGATCTCCCAAGCAACAGAGAAAGCACTGAGAGCTTCTGAGAAAACTGACTG 1920
DB 1861 TGAATCTGATCTCCCAAGCAACAGAGAAAGCACTGAGAGCTTCTGAGAAAACTGACTG 1918
QY 1921 CAATGCTCAGATGGAATGATCTTCAAGGCAAGTAAATTAATCACTCTGTGTAGAG 1980
DB 1921 CAATGCTCAGATGGAATGATCTTCAAGGCAAGTAAATTAATCACTCTGTGTAGAG 1978
QY 1981 ATACAGGTCAAGCTCAATTTCTTGAAGAACTGGATCAATTTGGAGAGCAGAGAAAG 2040
DB 1981 ATACAGGTCAAGCTCAATTTCTTGAAGAACTGGATCAATTTGGAGAGCAGAGAAAG 2038
QY 2041 AATTGGAAGAAAGAGAAATGTTACTTAAGGCAAGCAAGAGTGTCTTAATAAGAAAGATC 2100

Db 2039 ATTTGGAGAGAAAGAAATTTTACTTAAAGCAGCCAAAGTCTGCTTAATAAAGAAATC 2098
QY 2101 CAGAACAGCTGAGATTAAAGCAGAAAGCCAAAGATTACAGCAATTGGAATTCAGACAGA 2160
Db 2099 CAGAACAGCTGAGATTAAAGCAGAAAGCCAAAGATTACAGCAATTGGAATTCAGACAGA 2158
QY 2161 TACAGATAGAGACGCTAATCTCAGACGCTTCTCAGCTATTTGACCAAGAGAGAGAC 2220
Db 2159 TACAGATAGAGACGCTAATCTCAGACGCTTCTCAGCTATTTGACCAAGAGAGAGAC 2218
QY 2221 CACTAGAACCTGGAATTGAGGCGCTTAAAGACAACTTCTGCTTGGGACATCCAGCC 2280
Db 2219 CACTAGAACCTGGAATTGAGGCGCTTAAAGACAACTTCTGCTTGGGACATCCAGCC 2278
QY 2281 TGCAGACCCAGCAAGAGTTGACATGTCAGAAATCAAGAGATCAGAGATTCGCTAGGAGCTTGA 2340
Db 2279 TGCAGACCCAGCAAGAGTTGACATGTCAGAAATCAAGAGATCAGAGATTCGCTAGGAGCTTGA 2338
QY 2341 TATTTGTATGAGACAGAGAGGAGATGATTTCTGAGCTCTATACCTTGCCCTTC 2400
Db 2339 TATTTGTATGAGACAGAGGAGATGATTTCTGAGCTCTATACCTTGCCCTTC 2398
QY 2401 TGAAGTACACACTCCACTCTTCCATCCACATCCTGCTATTTACTGCGCAAGAGACACA 2460
Db 2399 TGAAGTACACACTCCACTCTTCCATCCACATCCTGCTATTTACTGCGCAAGAGACACA 2458
QY 2461 AAGCATTTGTCAGCTCTCTGAAATTTCAATTTCTGAAATTAACACACATGAAAGAG 2520
Db 2459 AAGCATTTGTCAGCTCTCTGAAATTTCAATTTCTGAAATTAACACACATGAAAGAG 2518
QY 2521 CATTTGTTACGATTAGACATTATTACCTTACCTAT 2558
Db 2519 CATTTGTTACGATTAGACATTATTACCTTACCTAT 2556

RESULT 2
HSTAF11
LOCUS HSTAF11 2556 bp mRNA linear PRI 14-NOV-1996
DEFINITION H.sapiens TAF1105 mRNA, partial.
ACCESSION Y09321
VERSION Y09321.1 GI:1669688
KEYWORDS TAF1105 gene.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2556)
AUTHORS Dikstein,R., Zhou,S. and Tjian,R.
TITLE Human TAF11 105 is a cell type-specific TFIID subunit related to
hTAF1130
JOURNAL Cell 87 (1), 137-146 (1996)
MEDLINE 97011146
PUBMED 8858156
REFERENCE 2 (bases 1 to 2556)
AUTHORS Dikstein,R.
TITLE Direct Submision
JOURNAL Submitted (07-NOV-1996) R. Dikstein, Weizmann Institute of Science,
Dept. Biochemistry, Rehovot, 76100, ISRAEL
FEATURES
source
1..2556
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HeLa"
1..2406
/gene="TAF1105"
1..2406
/gene="TAF1105"
/gene="TAF1105"
/codon_start=1
/product="7BP associated factor"
/protein_id="CAA70499.1"
/db_xref="GI:1669689"
/db_xref="SWISS-PROT:Q92750"

/translation="GTVTKVAVPAPKVSSEPLPAPQIVAKAPNTTITOPFANL
OLPQGVTLKNSGPIMLVSPQGVTPKRETSNITSRAVAPNTPTVICTVPSNSO
LIRKAVTPVKIKIOICTVTPKRETSNITSRAVAPNTPTVICTVPSNSO
LGASSTPENEKRNKNSAVOINISPMLEVKCNFLMLIKLACSGSOSPEMG
NKKLVEOLDAKLEAEFTKRYELKSSOPHLVPLKRSVAALRPLVSOTLPLA
OCVOOTSSDMVIACTTTPVTPVTPVSSSOSKSIIVSGATAPRTVSOTLPLA
GPVAKAGAVTLHSYGPAATGTTAGTGLTLOTSKPLVTVAANTVTSLOPEKPVY
GTAATLSLPAVTFEGSGAALCLPSVKRVFSCMDHCKPVGIPVOIKLADPGPVY
OPAGIPGSSSKOLFLPHVVOOSGSEKRYTISHSTLTLOKCGKTPVNTIIP
TSQPPASITLITOPKNTIISLOASPEFLDALOKRIDKIDITELNDVAVLIS
NEBACILITSELVGTLIQSKDEPFLDALOKRIDKIDITELNDVAVLIS
QATQERLGLLEKLTALAOHMTYKASENTILSDPSQLFLEKLDLQKORDE
EREMILKAERSKNEDEPOLRKLKAKAELOOLLELAOIHDAANTALAIQPRKRP
LESQTEGKMDLNLASGTSSTLATQOLHPRITRIRLIRLIFCMEREMKYSRALIYLA
LLK"

BASE COUNT 797 a 614 c 514 g 631 t
ORIGIN

Query Match 99.1%; Score 2534; DB 9; Length 2556;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2556; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 1 GGGACCCCTGTCACCAAAATGCTCGGTCAGCGCCCTCTTAAAGTACAGCGGCCCT 60
Db 1 GGGACCCCTGTCACCAAAATGCTCGGTCAGCGCCCTCTTAAAGTACAGCGGCCCT 60
QY 61 AGGCTGCGCTGCTCAGATAGTCGCGGTAAGGCCCCCAACACGCAATCCAGTTT 120
Db 61 AGGCTGCGCTGCTCAGATAGTCGCGGTAAGGCCCCCAACACGCAATCCAGTTT 120
QY 121 CCTGTAATTTGACAGCTTCTCCAGAGAACCGTTTGAATTAAGTAACAGTGGCTTG 180
Db 121 CCTGTAATTTGACAGCTTCTCCAGAGAACCGTTTGAATTAAGTAACAGTGGCTTG 180
QY 181 ATGTTGATTCCTCTCAGCAAACTGTACAGAGCCGAGACACAGTAATACATCTCA 240
Db 181 ATGTTGATTCCTCTCAGCAAACTGTACAGAGCCGAGACACAGTAATACATCTCA 240
QY 241 AGGCCAGAGTACAGAGCAATCTCAACAGTCAAACTGTACAGTGGCGCAACTTAC 300
Db 241 AGGCCAGAGTACAGAGCAATCTCAACAGTCAAACTGTACAGTGGCGCAACTTAC 300
QY 301 TCACAAATTAATCAAGAAAGTGCGACAGTCACTGTTAAATAATGACCAATAGGAAT 360
Db 301 TCACAAATTAATCAAGAAAGTGCGACAGTCACTGTTAAATAATGACCAATAGGAAT 360
QY 361 ACTGTTGTAACACGCTTCCGAAAGCGCTTCCAGTACATCTGTGCTGCCAACAGT 420
Db 361 ACTGTTGTAACACGCTTCCGAAAGCGCTTCCAGTACATCTGTGCTGCCAACAGT 420
QY 421 TGTCTCAGAGTACTCTCTGAAAGCAGTGAATCTGTAATCTCCGAAAGCGCTTCAAG 480
Db 421 TGTCTCAGAGTACTCTCTGAAAGCAGTGAATCTGTAATCTCCGAAAGCGCTTCAAG 480
QY 481 TTTGGGAGCAGTATCCCTTCAATAGAGCCCAATCTTAAAGCAGAGAACTCAGACG 540
Db 481 TTTGGGAGCAGTATCCCTTCAATAGAGCCCAATCTTAAAGCAGAGAACTCAGACG 540
QY 541 TGTAGATTAATCTTCTCCGACAACTGTGAAGAAATGTAAGAAATCAAGAACTCT 600
Db 541 TGTAGATTAATCTTCTCCGACAACTGTGAAGAAATGTAAGAAATCAAGAACTCT 600
QY 601 TGCATGTTAATAAAGTACAGATGAGTGAATCAAGTCCCTGAAATGGGCGCAAAATG 660
Db 601 TGCATGTTAATAAAGTACAGATGAGTGAATCAAGTCCCTGAAATGGGCGCAAAATG 660
QY 660 TGCATGTTAATAAAGTACAGATGAGTGAATCAAGTCCCTGAAATGGGCGCAAAATG 659
Db 660 TGCATGTTAATAAAGTACAGATGAGTGAATCAAGTCCCTGAAATGGGCGCAAAATG 659
QY 661 GAAGAAGCTGTGGAACAATTTTGAATGCAAAATGCAAGCAGAGAAATTTACTAGGAA 720
Db 661 GAAGAAGCTGTGGAACAATTTTGAATGCAAAATGCAAGCAGAGAAATTTACTAGGAA 720
QY 721 ACTGATTTGAACTCAAGTCTTCACTCAGCTCAGCTGCTCTTCTTAAAGAAAG 780
Db 721 ACTGATTTGAACTCAAGTCTTCACTCAGCTCAGCTGCTCTTCTTAAAGAAAG 779

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:59:29 ; Search time 21.8929 Seconds
(without alignments)
1145.042 Million cell updates/sec

Title: US-09-763-909-2

Perfect score: 4264
Sequence: 1 GLVTVKVAVSAPPKVSSGP.....KZHOHERALFTIRILLLITLY 852

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/2/1aa/5A.COMB.pep:*
4: /cgn2_6/prodata/2/1aa/5B.COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/prodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3990	93.6	801	1	US-08-725-012-2
2	1307.5	30.7	737	1	US-08-188-582-16
3	1307.5	30.7	737	1	US-08-646-715-16
4	775	18.2	921	1	US-08-188-582-2
5	775	18.2	921	1	US-08-646-715-2
6	237	5.6	2035	1	US-08-046-585-5
7	237	5.6	2035	1	US-08-393-703-5
8	237	5.6	2035	5	PCT-US93-11721-5
9	198	4.6	862	1	US-08-325-267A-4
10	197	4.6	2972	4	US-09-579-181-2
11	197	4.6	3118	4	US-09-579-181-1
12	191.5	4.5	894	3	US-08-362-525-22
13	191.5	4.5	894	3	US-08-971-692-15
14	184.5	4.3	1537	1	US-08-325-267A-2
15	178	4.2	1721	3	US-08-700-651-5
16	178	4.2	1721	3	US-08-928-361B-6
17	175.5	4.1	3569	4	US-08-061-376-5
18	172	4.0	750	4	US-09-165-239A-4
19	168	3.9	1837	3	US-08-928-361B-5
20	166	3.9	805	4	US-09-103-429A-4
21	164.5	3.9	752	1	US-08-244-189-2
22	164	3.8	903	2	US-08-853-310-2
23	160	3.8	1125	4	US-09-513-783A-152
24	160	3.8	1125	4	US-09-513-783A-22
25	158.5	3.7	786	4	US-09-103-429A-3
26	157.5	3.7	941	4	US-07-757-022B-14
27	157.5	3.7	1022	4	US-07-757-022B-84

28	157.5	3.7	1038	4	US-07-757-022B-74	Sequence 74, Appl
29	157.5	3.7	1049	4	US-07-757-022B-58	Sequence 58, Appl
30	157.5	3.7	1140	4	US-07-757-022B-104	Sequence 104, Appl
31	157.5	3.7	1270	4	US-07-757-022B-44	Sequence 44, Appl
32	157.5	3.7	1311	4	US-07-757-022B-42	Sequence 42, Appl
33	157.5	3.7	1313	4	US-07-757-022B-142	Sequence 142, Appl
34	157.5	3.7	1314	4	US-07-757-022B-50	Sequence 50, Appl
35	157.5	3.7	1320	4	US-07-757-022B-46	Sequence 46, Appl
36	157.5	3.7	1320	4	US-07-757-022B-60	Sequence 60, Appl
37	157.5	3.7	1354	4	US-07-757-022B-48	Sequence 48, Appl
38	157.5	3.7	1361	4	US-07-757-022B-40	Sequence 40, Appl
39	157.5	3.7	1363	4	US-07-757-022B-52	Sequence 52, Appl
40	157.5	3.7	1404	4	US-07-757-022B-2	Sequence 2, Appl
41	157.5	3.7	1404	4	US-07-757-022B-62	Sequence 62, Appl
42	157	3.7	2843	1	US-07-741-940-2	Sequence 2, Appl
43	157	3.7	2843	1	US-08-289-548A-2	Sequence 2, Appl
44	157	3.7	2843	1	US-08-452-654-2	Sequence 2, Appl
45	157	3.7	2843	2	US-08-370-235A-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-725-012-2
; Sequence 2, Application US/08725012
; Patent No. 5710025
; GENERAL INFORMATION:
; APPLICANT: Dikstein, Rivka
; APPLICANT: Tjian, Robert
; TITLE OF INVENTION: B-Cell Specific Transcription Factor
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESS: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/725,012
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B97-005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 801 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-725-012-2
Query Match 93.6% Score 3990; DB 1; Length 801;
Best Local Similarity 99.9%; Pred.No. 9,6e-319;
Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GLVTVKVAVSAPPKVSSGPLPAPQIVAVKAPNTTIOFPANTLOPPTVLIKSSGGL 60
DB 1 GLVTVKVAVSAPPKVSSGPLPAPQIVAVKAPNTTIOFPANTLOPPTVLIKSSGGL 60
QY 61 MAVSPQQTVAETVNTSITSRPAVPANPQTVKICVVPNSSQLIKKVAATPVKKLAQIGT 120

Db 61 MLVSPQQTVAETTSNITSRPVAVPAPNPOVTKICTVPSNSSQILKKVAVTPVKIKLAQIGT 120
 QY 121 TVTTPVPPSSVAVAPVPTVTPGKPLNTVTTLKSSLSGASSTPSNEPLKAENSA 180
 Db 121 TVTTPVPPSSVAVAPVPTVTPGKPLNTVTTLKSSLSGASSTPSNEPLKAENSA 180
 QY 181 VOINLSPMLBNVKKCNFLMLIKLACSGSSQSPMGONVKKVQLLDATIEEFPK 240
 Db 181 VOINLSPMLBNVKKCNFLMLIKLACSGSSQSPMGONVKKVQLLDATIEEFPK 240
 QY 241 LVYELKSSPOPHVLPFLKSVVALRQLLPNSQSFLOQCVOQTSDDMTATCTTTTSPV 300
 Db 241 LVYELKSSPOPHVLPFLKSVVALRQLLPNSQSFLOQCVOQTSDDMTATCTTTTSPV 300
 QY 301 VTTTSSSSQSEKSIIVSGATAPRVSVOTLPLAGPVGAKGVVTLHSVPTAATGTTA 360
 Db 301 VTTTSSSSQSEKSIIVSGATAPRVSVOTLPLAGPVGAKGVVTLHSVPTAATGTTA 360
 QY 361 GTGLQTSKPLVTSVANTVTVTSLOPKRPVSGTAVTSLSPATVEGSGAALCLPSVKP 420
 Db 361 GTGLQTSKPLVTSVANTVTVTSLOPKRPVSGTAVTSLSPATVEGSGAALCLPSVKP 420
 QY 421 VVSCPMHICKPVYGTVOVQIKLAPGVYLSOPACIPGSSSSKQLSLEFHVVOQPSGNEK 480
 Db 421 VVSCPMHICKPVYGTVOVQIKLAPGVYLSOPACIPGSSSSKQLSLEFHVVOQPSGNEK 480
 QY 481 QVTTTSSSTLTIOCKGOKTNPVTTIPTSOPASTLKQITLPKNLILSLQASPTOKR 540
 Db 481 QVTTTSSSTLTIOCKGOKTNPVTTIPTSOPASTLKQITLPKNLILSLQASPTOKR 540
 QY 541 IKENVTSCFDEDEDINDVTSMAGVNLNEENACILATNSEVLTIOCKDEPEFLTGALQ 600
 Db 541 IKENVTSCFDEDEDINDVTSMAGVNLNEENACILATNSEVLTIOCKDEPEFLTGALQ 600
 QY 601 KRLLDGGKHDITELNSDANVLSOATQERLGLLEKLAIAOHRMTTYKASBNYILCSD 660
 Db 601 KRLLDGGKHDITELNSDANVLSOATQERLGLLEKLAIAOHRMTTYKASBNYILCSD 660
 QY 661 TRSOLKFLKLDLEKORKEEREMILKAASRSNKEDDEOLRLQKAKELQQLAQT 720
 Db 661 TRSOLKFLKLDLEKORKEEREMILKAASRSNKEDDEOLRLQKAKELQQLAQT 720
 QY 721 QHRDANLITTAIGPRKKRPLESGIEGLKDNMLASGTSSLTATQOLRRPRTICRLDI 780
 Db 721 QHRDANLITTAIGPRKKRPLESGIEGLKDNMLASGTSSLTATQOLRRPRTICRLDI 780
 QY 781 FCMEOERKKYSRATYALALK 801
 Db 781 FCMEOERKKYSRATYALALK 801

RESULT 2
 US-08-188-582-16

Sequence 16, Application US/08188582
 Patent No. 5534410

GENERAL INFORMATION:

APPLICANT: Tjian, Robert
 APPLICANT: Comai, Lucio
 APPLICANT: Dynlacht, Brian D.
 APPLICANT: Hoey, Timothy
 APPLICANT: Ruppert, Siegfried
 APPLICANT: Tanese, Naoko
 APPLICANT: Wang, Edith
 APPLICANT: Weinzierl, Robert O.J.
 TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
 NUMBER OF SEQUENCES: 36
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
 STREET: 4 Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: USA

ZIP: 94111-4187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: PC-DOS/MS-DOS
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/188,582
 FILING DATE: 28-JAN-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Osman, Richard A
 REGISTRATION NUMBER: 36,627
 REFERENCE/DOCKET NUMBER: A-57650-2/AUT/RAO
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249
 TELETYPE: 910 277299
 INFORMATION FOR SEQ ID NO: 16:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 737, amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-188-582-16

Query Match 30.7%; Score 1307.5; DB 1; Length 737;
 Best Local Similarity 40.0%; Pred. No. 1,3e-98;
 Matches 330; Conservative 103; Mismatches 200; Indels 193; Gaps 22;

QY 13 PKRVSSG-----RLPAPQIVAAKAPNTTIOFPANLQIPGVLLKSNGLPLM 61
 Db 68 PTAATSGIRATLPPVYANRLPQ-----PONPNIQ--NFQLPQGVLVSENGQL 118
 QY 62 LVSPOQTVTR-----AETTSNITSRPVAVPAPNPOVTKICTVPSNSSQILKKVAVTPVKRL 115
 Db 119 MT-POQALQMOQAHAQOQTMAPRPATPTSAPOVISTVQAPGPIIAR-QVTP----- 172
 QY 116 AQIGTTVTVTKPPSSVAVAPVTSVVTTPGKPLNTYT--TLKSSLSGASS-----TPS 168
 Db 173 ----TTIIKQV---SOAQTTVPOSATLQSPGVQPOLVGAQAQTAAGTAVQGTGPQ 225
 QY 169 NE-PVLKAENSAVOINLSPMLBNVKKCNFLMLIKLACSGSSQSPMGONVKKVQLLD 227
 Db 226 RTVPGATTTSSAATP-----TMENVKCKNPLSTLILKASSGKOSTTAANVKELVONL 279
 QY 228 LDKATEEFPTRKLYVELKSSPOPHVLPFLKSVVALRQLLPNSQSFLOQCVOQ-----TS 283
 Db 280 LDGKTEADEFTSRLYRELINSSPOPLVLPFLKRSPLALROLTPDAAFTIQSQOQPPPTS 339
 QY 284 SDMVATCTTTTSPVTVTSVSSQSEKSIIVSGATAPRVSVOTLPLAGPVGAKGV 343
 Db 340 Q-----ATTALRAVAVLSSSVORTAGKTAATVTSALQPVLSL----- 376
 QY 344 VTLHSVGPATAAGTGTAGTGLLQTSKPLVTSVANTVTVTSLOPKRPVSGTAVTSLSPAV 403
 Db 377 ----TPTQVGVGKQGPPLVLIQ-----QPKRP-----GALIRPQV 410
 QY 404 TFEENSGAALPSVKRPVSCPMHICKPVYGTVOVQIKLAPGVYLSOPACIPGSSSKQ 463
 Db 411 TLTOT-----PVALRQPH-NRIMLTTPQIOL----- 437
 QY 464 LLSLFHVVOQPSGNEKQVTTTSSSTLTIOCKGOKTNPVTTIPTSOPASTLKQITL 523
 Db 438 -----NPLOPVPVKKPAVL 451
 QY 524 PGNKIL---SLQASPTOKRNIKENVTSCFDEDEDINDVTSMAGVNLNEENACILATSEL 580
 Db 452 PGTALSAVSAQAAAQKNLKEKGGSGFRDDINDVASMAGVNLSEESARILATSEL 511
 QY 581 VGTLLIOCKDEPEFLTGALQRLIDIGKHDITELNSDANVLSOATQERLGLLEKLA 640
 Db 512 VGTLLIOCKDEPEFLTGALQRLIDIGKHDITELNSDANVLSOATQERLGLLEKLA 640

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 20:54:50 ; Search time 16.4197 Seconds
(without alignments)
2152.161 Million cell updates/sec

Title: US-09-763-909-2

Perfect score: 4264
Sequence: 1 GTLVTKVAPVAPVAPVSGP.....KZHOERALTITLTLTLTY 852

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3990	93.6	801	T2DT_HUMAN	Q92750 homo sapien
2	1307.5	30.7	1083	T2D3_HUMAN	O00268 homo sapien
3	775	18.2	921	T2D3_DROME	P47825 drosophila
4	239.5	5.6	2035	HFCL_HUMAN	P51610 homo sapien
5	238	5.6	2090	HFCL_MESAU	P51611 mesocricetu
6	220	5.2	1367	AMYL_YEAST	P08640 saccharomyc
7	219	5.1	5179	MUC2_HUMAN	O02817 homo sapien
8	213.5	5.0	5376	ZAN_MOUSE	O08799 mus musculu
9	206.5	4.8	865	CPN_DROME	O02910 drosophila
10	205.5	4.8	1161	DAN4_YEAST	P47179 saccharomyc
11	203.5	4.8	1322	YAG3_YEAST	P39712 saccharomyc
12	201	4.7	725	AGAL_YEAST	P33233 saccharomyc
13	199	4.7	2700	ZAN_HUMAN	O9Y493 homo sapien
14	197.5	4.6	881	YUHE_YEAST	P47033 saccharomyc
15	197.5	4.6	1609	FTG2_YEAST	P25653 saccharomyc
16	193.5	4.5	670	VG50_HSV1	O00130 ictalunidi h
17	191.5	4.5	1260	ALST1_CANAL	P46300 candida alb
18	191	4.5	797	VGXL_HYVEB	P28968 equine herp
19	187.5	4.4	3726	TRX_DROME	P20659 drosophila
20	186	4.4	3178	TS89_CAEEL	O09624 caenorhabdi
21	184.5	4.3	1537	FLO1_YEAST	P32768 saccharomyc
22	182	4.3	1075	FLOS_YEAST	P38894 saccharomyc
23	182	4.3	3866	HRX_MOUSE	P35200 mus musculu
24	181.5	4.3	662	MUC1_XENILA	O05049 xenopus lae
25	181.5	4.3	1199	N121_RAT	P52591 rattus norv
26	181	4.2	1858	YK92_DICDI	O04893 saccharomyc
27	177.5	4.2	1140	YK96_YEAST	O9Y490 homo sapien
28	176	4.1	2541	TALI_HUMAN	P33636 homo sapien
29	175.5	4.1	2090	N214_HUMAN	O03164 homo sapien
30	175.5	4.1	3969	HRX_HUMAN	P41809 saccharomyc
31	173.5	4.1	1802	HRK1_YEAST	O99X47 mus musculu
32	173	4.1	2404	SON_MOUSE	O07284 epstein-bar
33	172	4.0	886	VGP3_EBVAB	

34	172	4.0	1018	1	HMW1_MYCPN	O50365 mycoplasma
35	172	4.0	1119	1	ALS3_CANAL	O74623 candida alb
36	172	4.0	1169	1	YK82_YEAST	P36170 saccharomyc
37	171.5	4.0	1780	1	YK26_CAEEL	P34333 caenorhabdi
38	170	4.0	1306	1	MSB2_YEAST	P34334 saccharomyc
39	169.5	4.0	388	1	MPY1_YEAST	P50105 saccharomyc
40	169.5	4.0	745	1	OCT1_PIG	O29076 sus scrofa
41	169.5	4.0	2426	1	SON_HUMAN	P18583 homo sapien
42	169	4.0	1229	1	N121_HUMAN	O9Y293 homo sapien
43	166.5	3.9	743	1	OCT1_HUMAN	P14659 homo sapien
44	165	3.9	606	1	SP2_HUMAN	O02086 homo sapien
45	165	3.9	2541	1	TALI_MOUSE	P26039 mus musculu

ALIGNMENTS

```

RESULT 1
T2DT_HUMAN          STANDARD;          PRT;          801 AA.
ID                  092750:
AC                  15-JUL-1998 (Rel. 36, Last sequence update)
DT                  15-JUL-1998 (Rel. 36, Last sequence update)
DE                  15-JUN-2002 (Rel. 41, Last annotation update)
DE                  Transcription Initiation Factor TFIID 105 kDa subunit (TAFII-105)
DE                  (TAFII105) (Fragment).
GN                  TAF4B OR TAF2C2 OR TAFII105.
OS                  Homo sapiens (Human).
OC                  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC                  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX                  NCBI_Taxid:9606;
RN                  [1]
RP                  MEDLINE=97011146; PubMed=8858156;
RX                  Dikstein R., Zhou S., Tjian R.;
RT                  "Human TAFII 105 is a cell type-specific TFIID subunit related to
RT                  hTAFII130."
RL                  Cell 87:137-146(1996).
CC                  -1- FUNCTION: CELL TYPE-SPECIFIC SUBUNIT OF TFIID THAT MAY FUNCTION AS
CC                  A GENE-SELECTIVE COACTIVATOR IN CERTAIN CELLS. TFIID IS A
CC                  MULTIMERIC PROTEIN COMPLEX THAT PLAYS A CENTRAL ROLE IN MEDIATING
CC                  PROMOTER RESPONSES TO VARIOUS ACTIVATORS AND REPRESSORS.
CC                  -1- SUBUNIT: TFIID IS COMPOSED OF RNA BINDING PROTEIN (TBP) AND A
CC                  NUMBER OF TBP-ASSOCIATED FACTORS (TAFs). TAFII105 MAY FORM
CC                  HETERODIMERS WITH TAFII130.
CC                  -1- SUBCELLULAR LOCATION: Nuclear.
CC                  -1- TISSUE SPECIFICITY: FOUND IN B CELLS BUT NOT IN OTHER TESTED
CC                  CELLS.
CC                  -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
CC                  -----
CC                  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC                  between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC                  the European Bioinformatics Institute. There are no restrictions on its
CC                  use by non-profit institutions as long as its content is in no way
CC                  modified and this statement is not removed. Usage by and for commercial
CC                  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC                  or send an email to license@sib-sib.ch).
CC                  -----
CC                  EMBL: Y09321; CAAT0499.1;
CC                  DR                  Genew; HGNC:11538; TAF4B.
CC                  DR                  MIM: 601689;
CC                  DR                  InterPro: IPR003894; TAF_hom.
CC                  DR                  SMART: SM00549; TAFH; 1.
CC                  KW                  Transcription regulation; Nuclear protein.
CC                  FT                  NON_TER
CC                  SQ                  SEQUENCE      801 AA; 8658 MW; D12B4932FEA9CD2 CRC64;
CC
Query Match          93.6%; Score 3990; DB 1; Length 801;
Best Local Similarity 99.9%; Pred. No. 5.1e-199;
Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
1 GTLVTKVAPVAPVSGPRLAPQIVYAKAPVTTTIOPRANLOPGTVLIKSNSGSL 60
|||||

```

Db 1 GTTATKAPKAPKPVSSGRLPAPQIVAKAPNTTIOFANLQIPCTVLKNSGRL 60
 QY 61 MLYSPQOQVTRAEITSNITSRAVPANPQVACICVPPNSSQLIKKAAVTPEVKAQIGT 120
 Db 61 MLYSPQOQVTRAEITSNITSRAVPANPQVACICVPPNSSQLIKKAAVTPEVKAQIGT 120
 QY 121 TTTTTPKSSVQSAVAPSVTVTPGKRLNVTTLKPSLGLASSPSEPNLKAENSA 180
 Db 121 TTTTTPKSSVQSAVAPSVTVTPGKRLNVTTLKPSLGLASSPSEPNLKAENSA 180
 QY 181 VOINLSPMLLENVKKCKNFNLAMLIKILACSGSSPEMGQVKKLVLPQILDAAKEAEFFTRK 240
 Db 181 VOINLSPMLLENVKKCKNFNLAMLIKILACSGSSPEMGQVKKLVLPQILDAAKEAEFFTRK 240
 QY 241 LVEELKSSPQPLVPLFKSVVALRQLLPNSQSFIOOCVQOQTSQDMVATCTTTVTSPV 300
 Db 241 LVEELKSSPQPLVPLFKSVVALRQLLPNSQSFIOOCVQOQTSQDMVATCTTTVTSPV 300
 QY 301 VTTTSSSSSEKSIIVSGATAPRTVSVQTLNPLAGPVKAKAVTLHSGPFAAAGTTA 360
 Db 301 VTTTSSSSSEKSIIVSGATAPRTVSVQTLNPLAGPVKAKAVTLHSGPFAAAGTTA 360
 QY 361 GTGLQTSKPLTVSVANVTVTLSLOPEKPVSGTAVTSLPVTGEGTGAICLPVYK 420
 Db 361 GTGLQTSKPLTVSVANVTVTLSLOPEKPVSGTAVTSLPVTGEGTGAICLPVYK 420
 QY 421 VVSPFMDHCKVIGTIPVOIKLAQEPVLSQAGIPPTGSSSKQLPSLFHVQOQPSGNEK 480
 Db 421 VVSPFMDHCKVIGTIPVOIKLAQEPVLSQAGIPPTGSSSKQLPSLFHVQOQPSGNEK 480
 QY 481 QVTTTSHSTLTIOKCGQGTMPVNTIPTSOPPASILKQILPCKTILSLQASPTOKNR 540
 Db 481 QVTTTSHSTLTIOKCGQGTMPVNTIPTSOPPASILKQILPCKTILSLQASPTOKNR 540
 QY 541 IKENVTSCFRDNDINDVTSAGVNLNENACILATNSLVTLLIQSCDEPFLFGLAQ 600
 Db 541 IKENVTSCFRDNDINDVTSAGVNLNENACILATNSLVTLLIQSCDEPFLFGLAQ 600
 QY 601 KRIDIGKHHITELNSDAVNLISQATQRLGLLEKLTALQHRMTYKASENTIILSD 660
 Db 601 KRIDIGKHHITELNSDAVNLISQATQRLGLLEKLTALQHRMTYKASENTIILSD 660
 QY 661 TFSQKFLFKLQLEKORDEEREMLKAAKRSNKEDPEQLRKOKAKELQOLELQOI 720
 Db 661 TFSQKFLFKLQLEKORDEEREMLKAAKRSNKEDPEQLRKOKAKELQOLELQOI 720
 QY 721 QHRDANLTATAAIGPRKKRPLESGIEGLKDNLLASGTSSTLTKQLHRRPRTICLRDLI 780
 Db 721 QHRDANLTATAAIGPRKKRPLESGIEGLKDNLLASGTSSTLTKQLHRRPRTICLRDLI 780
 QY 781 FCMEOEREMKYSRALYIALK 801
 Db 781 FCMEOEREMKYSRALYIALK 801

RESULT 2
 T2D3_HUMAN
 ID T2D3_HUMAN STANDARD: PRT: 1083 AA.
 AC 000268; Q99721; Q9BX42; Q9BR40;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Transcription initiation factor TFIID 135 kDa subunit (TAFII-135)
 GN (TAFII135) (TAFII-130) (TAFII130).
 OS TAFI OR TAF4A OR TAF2C1 OR TAF2C OR TAFII135 OR TAFII130.
 OC Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97336072; PubMed=9192867;
 RA Mengus G., May M., Carre L., Chambon P., Davidson I.;

RT "Human TAF(II)135 potentiates transcriptional activation by the AF-2s
 RT of the retinoic acid, vitamin D3, and thyroid hormone receptors in
 RT mammalian cells.";
 RL Genes Dev. 11:1381-1395(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.R., Bagguley C.L.,
 RA Bailey J., Barlow R.F., Bates K.N., Beard L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dham P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Grahame D.V., Griffiths C., Griffiths M.N.D., Gylliam R., Hall R.E.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jekoshts K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Leharvalho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McConachie L.J., McKay K., McMurtry A.A.,
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prichallingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkneen R., Sims S.,
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RT "The DNA sequence and comparative analysis of human chromosome 20.";
 RL Nature 414:865-871(2001).
 RN [3]
 RP SEQUENCE OF 105-1083 FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=97086442; PubMed=8942982;
 RA Taese N., Saluja D., Vassallo M.F., Chen J.-L., Admon A.;
 RA "Molecular cloning and analysis of two subunits of the human TFIID
 RT complex: hTAFII130 and hTAFII100.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:13611-13616(1996).
 CC -1- FUNCTION: MAKES PART OF TFIID IS A MULTIMERIC PROTEIN COMPLEX THAT
 CC PLAYS A CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS
 CC ACTIVATORS AND REPRESSORS. POTENTIATES TRANSCRIPTIONAL ACTIVATION
 CC BY THE AF-2S OF THE RETINOIC ACID, VITAMIN D3 AND THYROID HORMONE.
 CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
 CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Y11354; CAI2189.1; -
 CC EMBL: AL137077; CAC36006.1; -
 CC EMBL: AL109911; CAC2312.2; -
 CC EMBL: U75308; AAC50901.1; -
 CC TRANSFAC: T02328; -
 CC GeneW: HGNC:11537; TAF4.
 CC MIM: 601796; -
 CC InterPro: IPR003894; TAF_hom.
 CC SMART: SM00549; TAFH; 1.
 CC Transcription regulation: Nuclear protein.
 KW DOMAIN 39 42 POLY-HIS.
 FT DOMAIN 52 57 POLY-ALA.
 FT DOMAIN 98 101 POLY-GLY.
 FT DOMAIN 142 148 POLY-ALA.
 FT DOMAIN 268 275 POLY-PRO.

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2003, 22:29:40 : Search time 6475 Seconds

(Without alignments)
11497.294 Million cell updates/sec

Title: US-09-763-909-1

Sequence: 1 gggaccctgtgaccaaagt.....cttataactcttacctat 2558

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapept 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database:

GenBank: 1: gb_ba: 2: gb_ba: 3: gb_ba: 4: gb_ba: 5: gb_ba: 6: gb_ba: 7: gb_ba: 8: gb_ba: 9: gb_ba: 10: gb_ba: 11: gb_ba: 12: gb_ba: 13: gb_ba: 14: gb_ba: 15: gb_ba: 16: gb_ba: 17: gb_ba: 18: gb_ba: 19: gb_ba: 20: gb_ba: 21: gb_ba: 22: gb_ba: 23: gb_ba: 24: gb_ba: 25: gb_ba: 26: gb_ba: 27: gb_ba: 28: gb_ba: 29: gb_ba: 30: gb_ba: 31: gb_ba: 32: gb_ba: 33: gb_ba: 34: gb_ba: 35: gb_ba: 36: gb_ba: 37: gb_ba: 38: gb_ba: 39: gb_ba: 40: gb_ba: 41: gb_ba:

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2534	99.1	2556	6	181227
2	2534	99.1	2556	6	HSTAF11
3	1145.2	44.8	182805	2	AC017007
4	580	22.7	119081	2	AC121320
5	580	22.7	169240	9	AC022069
6	580	22.7	190782	2	AP002752
7	338.2	13.2	3603	6	123466
8	338.2	13.2	4233	9	HSU75308
9	337.2	13.2	3252	9	HSTAF113
10	325.8	12.7	2156	10	AY038601
11	294.6	11.5	174222	2	AP001197
12	294.6	11.5	175553	2	AP001096
13	294.6	11.5	182884	9	AC007996
14	294.6	11.5	190782	2	AP002752
15	148.6	5.8	333300	2	AC125091
16	148.6	5.8	16903	2	AC016839
17	148.6	5.8	61682	2	AC118057
18	147	5.7	277892	2	AC125105
19	146.6	5.7	205466	2	AC127767
20	143.8	5.6	33535	3	S63550
21	142.2	5.6	3745	3	AY069807
22	142.2	5.6	4615	3	DROTAFL10X
23	142.2	5.6	4615	6	123459
24	142.2	5.6	4615	6	145752
25	137.6	5.4	205466	2	AC127767
26	124.4	4.9	62974	2	AC118056
27	121.8	4.8	277892	2	AC125105
28	112.4	4.4	158509	2	AC114177
29	109.8	4.3	62974	2	AC118056
30	99	3.9	158509	2	AC114177
31	98.4	3.8	191918	2	AC102692
32	94.6	3.7	121982	9	HS107624
33	92	3.6	333300	2	AC125091
34	91.2	3.6	64035	2	AC119227
35	89.6	3.5	194426	2	AL663067
36	71	2.8	112277	2	AC113239
37	62.4	2.4	2147	9	AK097744
38	58.6	2.3	59665	2	AC014422
39	58.6	2.3	168479	3	AC093454
40	58.6	2.3	176056	3	AC010066
41	58.6	2.3	283821	3	AE003528
42	57.6	2.3	7218	6	166494
43	56.8	2.2	126712	9	AL137077
44	56.4	2.2	201470	2	AC113649
45	56.4	2.2	201470	2	AC113649

ALIGNMENTS

RESULT 1
LOCUS 181227 2556 bp DNA
DEFINITION Sequence 1 from patent US 5710025.
ACCESSION 181227
VERSION 181227.1 GI:3209517
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2556)
AUTHORS Dikstein,R. and Tjian,R.
TITLE Cell-type specific transcription factor
JOURNAL Patent: US 5710025-A 1 20-JAN-1998;
FEATURES Location/Qualifiers

source 1.2556
BASE COUNT 797 a 614 c 514 g 631 t
ORIGIN

Query Match 99.1%; Score 2534; DB 6; Length 2556;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2556; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 1 GGGACCCGTGTGACAAAGTGGTCCGGTCAGCGCCCTCTTAAAGTCAGACGCGCCCT 60
DB 1 GGGACCCGTGTGACAAAGTGGTCCGGTCAGCGCCCTCTTAAAGTCAGACGCGCCCT 60

QY 61 AGGCTGCTGCTCTCTGAGATAGTCCCGTAAAGCCGCCAACACGACGACATTCAGTTT 120
DB 61 AGGCTGCTGCTCTCTGAGATAGTCCCGTAAAGCCGCCAACACGACGACATTCAGTTT 120

QY 121 CCTGCTAATTTGACGCTTCTCCAGGAACGTTTGTATTAAGTAACAGTGGTCCGTTG 180
DB 121 CCTGCTAATTTGACGCTTCTCCAGGAACGTTTGTATTAAGTAACAGTGGTCCGTTG 180

QY 181 ATGTGTTATCTCTCTGACAAAGTGTAAAGAGCGAGACCAAGTAATACATACCTCA 240
DB 181 ATGTGTTATCTCTCTGACAAAGTGTAAAGAGCGAGACCAAGTAATACATACCTCA 240

QY 241 AGGCGACAGTACCAAGGATCTCTCAAAACAGTCAAAATCTGTACAGTGGCGAACTTAC 300
DB 241 AGGCGACAGTACCAAGGATCTCTCAAAACAGTCAAAATCTGTACAGTGGCGAACTTAC 300

QY 301 TCACAAATTAATCAAGAAAGTGGAGTACACGCTTTAAAAATTTGGCACAATAGGAACT 360
DB 301 TCACAAATTAATCAAGAAAGTGGAGTACACGCTTTAAAAATTTGGCACAATAGGAACT 360

QY 361 ACTGTGTAACACAGTGTCCGAAAGCCTTCTCAGTACATCTGTGCTGTGCCAACAGT 420
DB 361 ACTGTGTAACACAGTGTCCGAAAGCCTTCTCAGTACATCTGTGCTGTGCCAACAGT 420

QY 421 RGTGTCACAGTACACCTCCGAAAGCCATGTGATACGTACATACCTGAAAGCCTTCAAG 480
DB 421 -GTGTCACAGTACACCTCCGAAAGCCATGTGATACGTACATACCTGAAAGCCTTCAAG 480

QY 481 TTTGGAGAGTATCCACACTCTTCAATGAGCCCAATCTTAAAGCAGAGAACTCAGAGC 540
DB 481 TTTGGAGAGTATCCACACTCTTCAATGAGCCCAATCTTAAAGCAGAGAACTCAGAGC 540

QY 541 TGTTCAGATTAATCTTTCTCCGACATGTCTAGAAATGTGAAGAAATGCAAGAACTTCT 600
DB 541 TGTTCAGATTAATCTTTCTCCGACATGTCTAGAAATGTGAAGAAATGCAAGAACTTCT 600

QY 601 TCGAATGTTAATAAACTAGCATGTAGTATGATACAGTCCCTGGAATGGGGCAAAATGT 660
DB 601 TCGAATGTTAATAAACTAGCATGTAGTATGATACAGTCCCTGGAATGGGGCAAAATGT 660

QY 661 GAAGAAGCTGTGGAACAACTTTTGTGATGCAAAATGGAAGCAGAGAAATTTCTAGGAA 720
DB 661 GAAGAAGCTGTGGAACAACTTTTGTGATGCAAAATGGAAGCAGAGAAATTTCTAGGAA 720

QY 721 ACTGTATGTGAACCAAGTCTTACCTCAGCCTCAGCTGCTTCTTTCTTAAGAAAG 780
DB 721 ACTGTATGTGAACCAAGTCTTACCTCAGCCTCAGCTGCTTCTTTCTTAAGAAAG 780

QY 781 CGTGTGTGCTTACGACAACTTCTGCTTACCTCAGAGAGTTCATCAGCAATGTGTCA 840
DB 781 CGTGTGTGCTTACGACAACTTCTGCTTACCTCAGAGAGTTCATCAGCAATGTGTCA 840

QY 841 GCAGACTTCTAGTACATGTGTCTATCTTACCTTACCAAGAGTAACACTTCTCTGT 900
DB 841 GCAGACTTCTAGTACATGTGTCTATCTTACCTTACCAAGAGTAACACTTCTCTGT 900

QY 901 GGTGACAACTACAGTGTCTCAAGCCAGTCTGAAAGTCAATTAATTTCTGTGAGCAAC 960
DB 901 GGTGACAACTACAGTGTCTCAAGCCAGTCTGAAAGTCAATTAATTTCTGTGAGCAAC 960

QY 900 GGTGACAACTACAGTGTCTCAAGCCAGTCTGAAAGTCAATTAATTTCTGTGAGCAAC 959

QY 961 AGCACCAGAACTGTGTCACTGCAAACTTTGAACCCACTTGTCTGTCCAGTGGGACCAA 1020
DB 961 AGCACCAGAACTGTGTCACTGCAAACTTTGAACCCACTTGTCTGTCCAGTGGGACCAA 1020

QY 1021 AGCTGGAGTTGTGACACTTCAATCTGTGGGCCCCAAGTGGGCAAGAGGGAACAACAGC 1080
DB 1021 AGCTGGAGTTGTGACACTTCAATCTGTGGGCCCCAAGTGGGCAAGAGGGAACAACAGC 1080

QY 1081 TGGAACTGTGTTGCTTCAAGCTTCAAAACCACTTGTGACATCTGTGGCAAAACAGTGAC 1140
DB 1081 TGGAACTGTGTTGCTTCAAGCTTCAAAACCACTTGTGACATCTGTGGCAAAACAGTGAC 1140

QY 1141 CACGCTCTACCTCAACCTGAAAGCAGTGTCTGTGGAACGACAGTAACACTGTCTCT 1200
DB 1141 CACGCTCTACCTCAACCTGAAAGCAGTGTCTGTGGAACGACAGTAACACTGTCTCT 1200

QY 1201 TCCAGAGTAACTTTGAGAAACTCAGGTGACGATTTGTCTCCATCTGTGAAAC 1260
DB 1201 TCCAGAGTAACTTTGAGAAACTCAGGTGACGATTTGTCTCCATCTGTGAAAC 1260

QY 1261 TGTGTTTCTCTCTGCTGGGACCACTGTGCAAGCCTGTATTTGGGACTCCAGTTCAAT 1320
DB 1261 TGTGTTTCTCTCTGCTGGGACCACTGTGCAAGCCTGTATTTGGGACTCCAGTTCAAT 1320

QY 1321 CAAACTTGGCCAGCGGGCCCTGTCTTTCACAACAGCTGGGATTCACAAGCAGTTTC 1380
DB 1321 CAAACTTGGCCAGCGGGCCCTGTCTTTCACAACAGCTGGGATTCACAAGCAGTTTC 1380

QY 1381 AAGCAGCAACTTCTCATTTGTTTCAAGTGTGACAGGCTTCAAGAGGCAATGAAA 1440
DB 1381 AAGCAGCAACTTCTCATTTGTTTCAAGTGTGACAGGCTTCAAGAGGCAATGAAA 1440

QY 1441 ACAAGTACCAATTTCACTATCTCAACATTTGACCATTCAGAAATGTGAGAGAGAAC 1500
DB 1441 ACAAGTACCAATTTCACTATCTCAACATTTGACCATTCAGAAATGTGAGAGAGAAC 1500

QY 1501 GATGCCAGTGAACACCAATTAATCTAGTACAGTTTCTCCAGCTTCCATTTGTAAGCA 1560
DB 1501 GATGCCAGTGAACACCAATTAATCTAGTACAGTTTCTCCAGCTTCCATTTGTAAGCA 1560

QY 1561 AATTACCTGCTGGAATTAATTAATCTGCACTTCAAGCATCTCTACAGAAATA 1620
DB 1561 AATTACCTGCTGGAATTAATTAATCTGCACTTCAAGCATCTCTACAGAAATA 1620

QY 1621 GAATTAAGAGATGTAACTATCTGCTCCGAGATGAGATGACATCAATGATGTGACTT 1680
DB 1621 GAATTAAGAGATGTAACTATCTGCTCCGAGATGAGATGACATCAATGATGTGACTT 1680

QY 1681 CTATGCGAGGGGTCAACCTTAATGAAGAAATGCTGTGATCTTGAACAACACTGTAAT 1740
DB 1681 CTATGCGAGGGGTCAACCTTAATGAAGAAATGCTGTGATCTTGAACAACACTGTAAT 1740

QY 1741 TGTGTCGACACTCATTCAGTCACTGTAAGATGAACTTCTTTTATTTGGAGCTCTAC 1800
DB 1741 TGTGTCGACACTCATTCAGTCACTGTAAGATGAACTTCTTTTATTTGGAGCTCTAC 1800

QY 1801 AAAAGAAATCTTAGACATTTGTTAAAGCATGACATTAAGAACTTAATCTGATGCTG 1860
DB 1801 AAAAGAAATCTTAGACATTTGTTAAAGCATGACATTAAGAACTTAATCTGATGCTG 1860

QY 1861 TGAACCTGATCTCCCAAGCAACACAGAGAGCTACGAGGCTTCTTGAAGAAACAGCTG 1920
DB 1861 TGAACCTGATCTCCCAAGCAACACAGAGAGCTACGAGGCTTCTTGAAGAAACAGCTG 1920

QY 1921 CAATTCCTCAGATCAATGATGATTAAGCAAGCAAGTGAATTTCAATCTGTGATG 1980
DB 1921 CAATTCCTCAGATCAATGATGATTAAGCAAGCAAGTGAATTTCAATCTGTGATG 1980

QY 1981 ATACAGGTTCACAGCTCAAAATTTCTTGAAGAGCTGATCAATTTGAGAGACAGAGAAAG 2040
DB 1981 ATACAGGTTCACAGCTCAAAATTTCTTGAAGAGCTGATCAATTTGAGAGACAGAGAAAG 2040

QY 2041 AATTGGAAGAAAGAAATGTTACTTAAGGACCCAGAGTCTCTTAAATTAAGAAAGATC 2100


```

gene      1..2406
          /gene="TAIR1105"
CDS       <1..2406
          /gene="TAIR1105"
          /codon_start=1
          /product="7BP associated factor"
          /protein_id="CAI04049.1"
          /db_xref="GI:1696689"
          /db_xref="SWISS-PROT:Q92750"

```

Query Match	99.1%	Score 2534	DB 9	Length 2556
Best Local Similarity	99.9%	Pred. No. 0	Mismatches 0	Indels 2
Matches 2556	Conservative 0	Indels 0	Indels 0	Gaps 2
QY 1	GGGACCCGTGGACCAAAAGTGCTCCGGGTGAGGCGCCCTCTTAAGTACGAGAGGGCCCT	60		
Db 1	GGGACCCGTGGACCAAAAGTGCTCCGGGTGAGGCGCCCTCTTAAGTACGAGAGGGCCCT	60		
QY 61	AGGCTCGCTGCTCCTCAGATAGTGGCGGTGAAAGGCCCAACACACAGCAATCCAGTTT	120		
Db 61	AGGCTCGCTGCTCCTCAGATAGTGGCGGTGAAAGGCCCAACACACAGCAATCCAGTTT	120		
QY 121	CCTGCTAATTGGACGCTTCCCTCCAGAAACCGTTTGATTAAGTAAACAGTGGTCCGTG	180		
Db 121	CCTGCTAATTGGACGCTTCCCTCCAGAAACCGTTTGATTAAGTAAACAGTGGTCCGTG	180		
QY 181	ATGTGGATCTCCTCCAGCAACGTGTAAACAGAGCCGAGACACAGTAATATAACCTCA	240		
Db 181	ATGTGGATCTCCTCCAGCAACGTGTAAACAGAGCCGAGACACAGTAATATAACCTCA	240		
QY 241	AGGCCAGCAGTACGAGGAAATCCTCAAAACAGTCAAAATCTGTACAGTCCCAACTTAC	300		
Db 241	AGGCCAGCAGTACGAGGAAATCCTCAAAACAGTCAAAATCTGTACAGTCCCAACTTAC	300		
QY 301	TCCAATTAATCAAGAAAGGCGCAGTACACCTGTATAAAATTTGGCACAATATGAACT	360		
Db 301	TCCAATTAATCAAGAAAGGCGCAGTACACCTGTATAAAATTTGGCACAATATGAACT	360		
QY 361	ACTGTGGTAAACCACTGTGTCCGAAAGCCTTCCCTCACTAACAATGTGGGCGTCCAAACAGT	420		
Db 361	ACTGTGGTAAACCACTGTGTCCGAAAGCCTTCCCTCACTAACAATGTGGGCGTCCAAACAGT	420		
QY 421	RGTTGTCACAAGTACCTCTCGAAAGCCATTGAATATCTGTAACTACCTCGAAGCCTTCAG	480		
Db 421	RGTTGTCACAAGTACCTCTCGAAAGCCATTGAATATCTGTAACTACCTCGAAGCCTTCAG	480		
QY 481	TTTGGAGCAATCACTCACTCCTCAATTAAGGCCCAACTTAAGCAGAGAACTACAGC	540		
Db 481	TTTGGAGCAATCACTCACTCCTCAATTAAGGCCCAACTTAAGCAGAGAACTACAGC	540		
QY 541	TGTTCAATTAATCTTTCCTCGAACAATGTAGAAAATGTAGAAAATGTAGAAACTTCT	600		
Db 541	TGTTCAATTAATCTTTCCTCGAACAATGTAGAAAATGTAGAAAATGTAGAAACTTCT	600		
QY 601	TGCAATTTAATAAACTAGCATGTAGTGAATCAAGTCCCTGAAATGGGGCAAAATGT	660		
Db 601	TGCAATTTAATAAACTAGCATGTAGTGAATCAAGTCCCTGAAATGGGGCAAAATGT	660		
QY 661	GAAGAACCTGTGGAAACAATTTTGGATGGAATAAATCGAAGCAAGAAATTTACTAGAA	720		
Db 661	GAAGAACCTGTGGAAACAATTTTGGATGGAATAAATCGAAGCAAGAAATTTACTAGAA	720		
QY 721	ACTGTATGTGAACCAAGTCTTCACTCAGCCTCACTGTTCTTTTCTTAAGAAAG	780		
Db 721	ACTGTATGTGAACCAAGTCTTCACTCAGCCTCACTGTTCTTTTCTTAAGAAAG	780		

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:59:29 ; Search time 21.8929 Seconds
(without alignments)
1145.042 Million cell updates/sec

Title: US-09-763-909-2

Perfect score: 4264
Sequence: 1 GTLVTKVAPVAPKVSNGP.....KZHOHERALFTITLTLTY 852

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3990	93.6	801	1 US-08-725-012-2	Sequence 2, Appli
2	1307.5	30.7	737	1 US-08-188-582-16	Sequence 16, Appli
3	1307.5	30.7	737	1 US-08-646-715-16	Sequence 16, Appli
4	775	18.2	921	1 US-08-188-582-2	Sequence 2, Appli
5	775	18.2	921	1 US-08-646-715-2	Sequence 2, Appli
6	237	5.6	2035	1 US-08-046-585-5	Sequence 5, Appli
7	237	5.6	2035	1 US-08-393-703-5	Sequence 5, Appli
8	237	5.6	2035	5 PCT-US93-11721-5	Sequence 5, Appli
9	198	4.6	862	1 US-08-325-267A-4	Sequence 4, Appli
10	197	4.6	2972	1 US-08-325-267A-4	Sequence 4, Appli
11	197	4.6	3118	4 US-09-579-181-2	Sequence 2, Appli
12	191.5	4.5	894	3 US-08-362-525-22	Sequence 2, Appli
13	191.5	4.5	894	3 US-08-971-692-15	Sequence 22, Appli
14	184.5	4.3	1537	1 US-08-325-267A-2	Sequence 15, Appli
15	178	4.2	1721	3 US-08-700-651-5	Sequence 2, Appli
16	178	4.2	1721	3 US-08-928-361B-6	Sequence 6, Appli
17	175.5	4.1	3969	4 US-08-061-316-5	Sequence 5, Appli
18	172	4.0	750	4 US-09-165-239A-4	Sequence 5, Appli
19	168	3.9	1837	3 US-08-928-361B-5	Sequence 5, Appli
20	166	3.9	805	4 US-09-103-429A-4	Sequence 4, Appli
21	164.5	3.9	752	1 US-08-244-189-2	Sequence 2, Appli
22	164	3.8	903	2 US-08-853-310-2	Sequence 2, Appli
23	160	3.8	1125	4 US-09-513-783A-152	Sequence 12, Appli
24	160	3.8	1610	4 US-09-513-783A-22	Sequence 12, Appli
25	158.5	3.7	786	4 US-09-103-429A-3	Sequence 3, Appli
26	157.5	3.7	941	4 US-07-757-022B-14	Sequence 14, Appli
27	157.5	3.7	1022	4 US-07-757-022B-84	Sequence 84, Appli

28	157.5	3.7	1038	4 US-07-757-022B-74	Sequence 74, Appli
29	157.5	3.7	1049	4 US-07-757-022B-58	Sequence 58, Appli
30	157.5	3.7	1140	4 US-07-757-022B-104	Sequence 104, App
31	157.5	3.7	1270	4 US-07-757-022B-44	Sequence 44, Appli
32	157.5	3.7	1311	4 US-07-757-022B-42	Sequence 42, Appli
33	157.5	3.7	1313	4 US-07-757-022B-142	Sequence 142, App
34	157.5	3.7	1314	4 US-07-757-022B-50	Sequence 50, Appli
35	157.5	3.7	1320	4 US-07-757-022B-46	Sequence 46, Appli
36	157.5	3.7	1354	4 US-07-757-022B-60	Sequence 60, Appli
37	157.5	3.7	1361	4 US-07-757-022B-48	Sequence 48, Appli
38	157.5	3.7	1361	4 US-07-757-022B-40	Sequence 40, Appli
39	157.5	3.7	1363	4 US-07-757-022B-52	Sequence 52, Appli
40	157.5	3.7	1404	4 US-07-757-022B-2	Sequence 2, Appli
41	157.5	3.7	1404	4 US-07-757-022B-62	Sequence 62, Appli
42	157	3.7	2843	1 US-07-741-940-2	Sequence 2, Appli
43	157	3.7	2843	1 US-08-289-548A-2	Sequence 2, Appli
44	157	3.7	2843	1 US-08-452-654-2	Sequence 2, Appli
45	157	3.7	2843	2 US-08-370-235A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-725-012-2
Sequence 2, Application US/08725012
Patent No. 5710025

GENERAL INFORMATION:

APPLICANT: Dikstein, Rivka
APPLICANT: Tjian, Robert
TITLE OF INVENTION: B-Cell Specific Transcription Factor
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/725,012
FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: B97-005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 801 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-725-012-2

Query Match 93.6% Score 3990: DB 1: Length 801:
Best Local Similarity 99.9% Pred. No. 9.6e-319;
Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTLVTKVAPVAPKVSNGPRLPAPOIYAVAKAPMTTITOPANOLPQGVILIKNSGPL 60
DB 1 GTLVTKVAPVAPKVSNGPRLPAPOIYAVAKAPMTTITOPANOLPQGVILIKNSGPL 60
QY 61 MLVSPQQTVAETTSNTTSPPAVANPQYVICTVPNSSQLIKKAVTPVKRLAQTCT 120
|||||

```

Db      61  MLYSPQOTVTRAEITSNITSRPAVPANPQVYKICTVPSNSSLIKRYAVTPVKLAQIGT 120
QY      121  TVVTVPKPSSVOSVAVPTSVTVTPGKPIINTVTTTLKPSLSGASSTSPNPNKAESAA 180
Db      121  TVVTVPKPSSVOSVAVPTSVTVTPGKPIINTVTTTLKPSLSGASSTSPNPNKAESAA 180
QY      181  VOINLSPTMLENWKCKNFIAMLIKACSSQSPSEMGOVKKLVEQLLDKAEFEPTK 240
Db      181  VOINLSPTMLENWKCKNFIAMLIKACSSQSPSEMGOVKKLVEQLLDKAEFEPTK 240
QY      241  LVELEKSSPOPLVPLFKKSVVALROLPLPSOSFIOQCVOQTSDBVIATCTTVTTPV 300
Db      241  LVELEKSSPOPLVPLFKKSVVALROLPLPSOSFIOQCVOQTSDBVIATCTTVTTPV 300
QY      301  VTTVSSSOSEKSIIVSGAAPTAVSQTUNPLAGPVGAAGVVTLSHVSPTAATGTTA 360
Db      301  VTTVSSSOSEKSIIVSGAAPTAVSQTUNPLAGPVGAAGVVTLSHVSPTAATGTTA 360
QY      361  GTGLQTSKPLVTSVANTVTTVTSLOPEKPVSGTAVTSLPVTFGTSGAALCPSVPR 420
Db      361  GTGLQTSKPLVTSVANTVTTVTSLOPEKPVSGTAVTSLPVTFGTSGAALCPSVPR 420
QY      421  VVSFCMDHICKPVIGTPVOIKLAOPBPVLSOPAGIPTGSSSKOLFSLFHVVOQPSGNEK 480
Db      421  VVSFCMDHICKPVIGTPVOIKLAOPBPVLSOPAGIPTGSSSKOLFSLFHVVOQPSGNEK 480
QY      481  QVTTSHSSTLTLOKCGOKTNPVNTIPTSOPFPASITLQITLPKNTLISLASPTOKNR 540
Db      481  QVTTSHSSTLTLOKCGOKTNPVNTIPTSOPFPASITLQITLPKNTLISLASPTOKNR 540
QY      541  IKENTTSCRRDDINDVTSMAGVNLNENACILATNSELVGTLLIOSCKDEPFLTGALO 600
Db      541  IKENTTSCRRDDINDVTSMAGVNLNENACILATNSELVGTLLIOSCKDEPFLTGALO 600
QY      601  KRILDIGKKHDTLNSDAVNLISQATQERLGLLEKLTALQOHMTTYKASENTILCSO 660
Db      601  KRILDIGKKHDTLNSDAVNLISQATQERLGLLEKLTALQOHMTTYKASENTILCSO 660
QY      661  TRSOLKFEKLDQLEKORRDLEREMELKAAKSRNKEDPEQLRKOKAKELQOUELAQI 720
Db      661  TRSOLKFEKLDQLEKORRDLEREMELKAAKSRNKEDPEQLRKOKAKELQOUELAQI 720
QY      721  QHRDANLTATAIGRRKRPLESGIEGKDNILASGSSSLATKQALHPRITRILCRLDI 780
Db      721  QHRDANLTATAIGRRKRPLESGIEGKDNILASGSSSLATKQALHPRITRILCRLDI 780
QY      781  FCMOEREMKYSRALYIALLK 801
Db      781  FCMOEREMKYSRALYIALLK 801

```

RESULT 2

US-08-188-582-16
Sequence 16, Application us/08188582
Patent No. 5534410

GENERAL INFORMATION:

APPLICANT: Tjian, Robert
APPLICANT: Comai, Lucio
APPLICANT: Dynlacht, Brian D.
APPLICANT: Hoey, Timothy
APPLICANT: Ruppert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
NUMBER OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
CORRESPONDENCE ADDRESS: 36
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA

```

ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/188,582
FILING DATE: 28-JAN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RMO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 737 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-188-582-16

```

Query Match 30.7%; Score 1307.5; DB 1; Length 737;
Best Local Similarity 40.0%; Pred. No. 1,3e-98;
Matches 330; Conservative 103; Mismatches 200; Indels 193; Gaps 22;

```

QY      13  PPKVSSG-----PRLPAPQIVAKAPNTTIOFPANLQLPBGTVLKKSSGLM 61
Db      68  PTATSGIRATLTPVPLRLPQ-----PQNPINQ---NFQLPQMLVLRSSNGQL 118
QY      62  LVSPOQVTR-----ATTSNITSRPAVPANPQTVKICTVPSNSSLIKRYAVTPVKL 115
Db      119  MI-PQALAOQAOAHAPQPTTMAPRPATPTSAAPVQISTVQAPQPTIAR-QVTP----- 172
QY      116  AQIGTVTVTPKPSVOSVAVPTSVTVTPGKPLNTYT--TLKPSLSGASS-----TPS 168
Db      173  ----TTIKQV---SQAOITVQPSATLQKSPGVOPQLVGGAAQTASIGTAVQGTGPQ 225
QY      169  NE-PRLAENSAVOINLSPTMLENWKCKNFIAMLIKACSSQSPSEMGOVKKLVEQL 227
Db      226  RTVPGATTTSSAATE-----TMENVKCKNFIETLILKIASGSGQSTETANVELVONL 279
QY      228  LDAKIEAEFEPRKLYVELKSSPOPLVPLFKKSVVALROLPLPSOSFIOQCVOQ-----TS 283
Db      280  LDGKIEADEFTSRILYRELNSSPOPLVPLFKKSLPALROLTPDSAPRTQSQQPPPTPS 339
QY      284  SDNVIACTCTTCTVTTSPVTTVSSSOSEKSIIVSGAAPTAVSQTUNPLAGPVGAAGV 343
Db      340  Q-----ATTALAVLVSSVQRTACKTAAYTSALQPVLSL----- 376
QY      344  VTLHSGVPTAATGGTAGTGLQTSKPLVTSVANTVTTVTSLOPEKPVSGTAVTSLP 403
Db      377  -----TPPTGVGVCKQCGQPPPLVQ-----QPPKP-----GALIRPPQV 410
QY      404  TFGTSGAALCPSVPRVVSFCMDHICKPVIGTPVOIKLAOPBPVLSOPAGIPTGSSSKO 463
Db      411  TLVQT-----PMVALRQPH-NRIMLTTPQOIQ----- 437
QY      464  LPSLFHVVOQPSGNEKQVTTTSHSSTLTLOKCGOKTNPVNTIPTSOPFPASITLQITL 533
Db      438  ----- 541
QY      524  PGKNTL-----SLQASPTQKRIKENYTCFRDEDDINDVTSMAGVNLNENACILATNSEL 580
Db      452  PGTKALSAVSAQAAQAAQKKKIKKPGGSPRDDDDINDVTSMAGVNLNENACILATNSEL 511
QY      581  VGTLLIOSCKDEPFLTGALOKRLDIGKKHDTLNSDAVNLISQATQERLGLLEKLTAL 640
Db      512  VGTLLRSCDEFTLLQAPLQRLILEIGKKHGTETLHPDVVSVYSHATQORLQNTVKEISE 571

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 20:54:50 ; Search time 16.4197 Seconds
(without alignments)
2152.161 Million cell updates/sec

Title: US-09-763-909-2

Perfect score: 4264
Sequence: 1 GTLVTKVAPVAPRYSSGP.....KZHOHERALFTIRLLTLTV 852

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3990	93.6	801	T2DT_HUMAN	Q92750 homo sapien
2	1307.5	30.7	1083	T2D3_HUMAN	000268 homo sapien
3	775	18.2	921	T2D3_DROME	P47825 drosophila
4	239.5	5.6	2035	HFCL_HUMAN	P51610 homo sapien
5	238	5.6	2090	HFCL_MESAU	P51611 mesocricetu
6	220	5.2	1367	AMTH_YEAST	P08640 saccharomyc
7	219	5.1	5179	MUC2_HUMAN	002817 homo sapien
8	213.5	5.0	5376	ZAN_MOUSE	088799 mus musculu
9	206.5	4.8	865	CPN_DROME	002910 drosophila
10	205.5	4.8	1161	DAN4_YEAST	P47179 saccharomyc
11	203.5	4.8	1322	YAG3_YEAST	P39712 saccharomyc
12	201	4.7	725	AGAL_YEAST	P32323 saccharomyc
13	199	4.7	2700	ZAN_HUMAN	094493 homo sapien
14	197.5	4.6	881	YJH8_YEAST	P47033 saccharomyc
15	197.5	4.6	1609	FTG2_YEAST	P25653 saccharomyc
16	193.5	4.5	670	VG50_HSV1	000130 ictaluriid b
17	191.5	4.5	1260	ALSL_CANAL	P46590 candida alb
18	191	4.5	797	VGIX_HAYEB	P28968 equine hepr
19	187.5	4.4	3726	TRX_DROME	P20659 drosophila
20	186	4.4	3178	VS89_DROME	P09624 caenorhabdi
21	184.5	4.3	1537	FLOI_YEAST	P32768 saccharomyc
22	182	4.3	1075	FLO5_YEAST	P38894 saccharomyc
23	182	4.3	3866	HRX_MOUSE	P55200 mus musculu
24	181.5	4.3	662	MUC1_XENLA	005049 xenopus lae
25	181.5	4.3	1199	N121_RAT	P52591 rattus norv
26	181	4.2	1858	P3K2_DICDI	P54674 dictyosteli
27	177.5	4.2	1140	YH96_YEAST	004893 saccharomyc
28	176	4.1	2541	TAL1_HUMAN	094490 homo sapien
29	175.5	4.1	2090	N214_HUMAN	P33658 homo sapien
30	175.5	4.1	3969	HR1_HUMAN	003164 homo sapien
31	173.5	4.1	1802	HR1_YEAST	P41809 saccharomyc
32	173	4.0	2404	SON_MOUSE	09447 mus musculu
33	172	4.0	886	VGP3_EBVA8	007284 epstein-bar

34	172	4.0	1018	1	HMW1_MYCPN	Q50365 mycoplasma
35	172	4.0	1119	1	AL53_CANAL	074623 candida alb
36	172	4.0	1169	1	YK82_YEAST	P36170 saccharomyc
37	171.5	4.0	1780	1	YK26_CAEEL	P34333 caenorhabdi
38	170	4.0	1306	1	MSB2_YEAST	P32334 saccharomyc
39	169.5	4.0	388	1	MPRT_YEAST	P50105 saccharomyc
40	169.5	4.0	745	1	OCT1_PIG	Q29076 sus scrofa
41	169.5	4.0	2426	1	SON_HUMAN	P18583 homo sapien
42	169	4.0	1229	1	N121_HUMAN	094293 homo sapien
43	166.5	3.9	743	1	OCT1_HUMAN	P14859 homo sapien
44	165	3.9	606	1	SP2_HUMAN	002086 homo sapien
45	165	3.9	2541	1	TAL1_MOUSE	P26039 mus musculu

ALIGNMENTS

```

RESULT 1
ID T2DT_HUMAN STANDARD: PRT: 801 AA.
AC Q92750:
DF 15-JUL-1998 (Rel. 36, Created)
DF 15-JUL-1998 (Rel. 36, Last sequence update)
DF 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcription Initiation factor TFIID 105 kDa subunit (TAFII-105)
DE (TAFII105) (Fragment).
DE TAFAB OR TAF2C2 OR TAFII105.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP MEDLINE=97011146; PubMed=8858156;
RX Dikstein R., Zhou S., Tjian R.;
RT "Human TAFII 105 is a cell type-specific TFIID subunit related to
   hTAFII130."
RL Cell 87:137-146(1996).
CC -1- FUNCTION: CELL TYPE-SPECIFIC SUBUNIT OF TFIID THAT MAY FUNCTION AS
CC A GENE-SELECTIVE COACTIVATOR IN CERTAIN CELLS. TFIID IS A
CC MULTIMERIC PROTEIN COMPLEX THAT PLAYS A CENTRAL ROLE IN MEDIATING
CC PROMOTER RESPONSES TO VARIOUS ACTIVATORS AND REPRESSORS.
CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs). TAFII105 MAY FORM
CC HETERODIMERS WITH TAFII130.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: FOUND IN B CELLS BUT NOT IN OTHER TESTED
CC CELLS.
CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL: Y09321; CAAT70499.1;
CC Genew: HGNC:11538; TAFAB.
CC MIM: 601689;
CC DR InterPro: IPR003894; TAF_hom.
CC DR SMART: SM00549; TAFH; 1.
CC KW Transcription regulation; Nuclear protein.
CC FT NON_TER
CC SQ SEQUENCE 801 AA; 85558 MW; D12B4932EFA9CD2 CRC64;

Query Match 93.6%; Score 3990; DB 1; Length 801;
Best Local Similarity 99.9%; Pred. No. 5.1e-199;
Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
1 GTLVTKVAPVAPRYSSGPRAPQIVAVKAPNTTIOFPANLQLPCTVLIKNSGPT 60
|||||

```

Db 1 GTLVTKAVPAPVAPKSSGRLPAPQIVAVKAPNTTIIQFPANQLPFGTVLISNSGPL 60
 QY 61 MLYSPQOTVTRAEFTSNITSRAVPANPQVYKICTVNSSQLIKKVAVTPVKRLAQIGT 120
 Db 61 MLYSPQOTVTRAEFTSNITSRAVPANPQVYKICTVNSSQLIKKVAVTPVKRLAQIGT 120
 QY 121 TVYTVTRKPSVSVAVPTVYVYTPKRPPLNTVYTLKPSISGASTSNENKLAENSA 180
 Db 121 TVYTVTRKPSVSVAVPTVYVYTPKRPPLNTVYTLKPSISGASTSNENKLAENSA 180
 QY 181 VOINLSPMLENKKCNFLAMLKIKACSGSGSPGEMGNVYKYLEQOLLDARIEAEFTRK 240
 Db 181 VOINLSPMLENKKCNFLAMLKIKACSGSGSPGEMGNVYKYLEQOLLDARIEAEFTRK 240
 QY 241 LVEELKSPQPHLPVPELKKSVVALROLNPSNSQFIOOCVOYSSDMYIATCTTVTSPV 300
 Db 241 LVEELKSPQPHLPVPELKKSVVALROLNPSNSQFIOOCVOYSSDMYIATCTTVTSPV 300
 QY 301 VTTVSSSSQSEKSIIVSGATAPRVSVQTLNPLAGPAGARAGVYTLHSVGTATGTTA 360
 Db 301 VTTVSSSSQSEKSIIVSGATAPRVSVQTLNPLAGPAGARAGVYTLHSVGTATGTTA 360
 QY 361 GTGLQTSKPLVTSVANTVTVTSIQPEKPVVSGTAVTLSPAVTFGETSGAICLPVYK 420
 Db 361 GTGLQTSKPLVTSVANTVTVTSIQPEKPVVSGTAVTLSPAVTFGETSGAICLPVYK 420
 QY 421 VVSFCMDHCKPVYIGTPVQIKLAOPPVLSOPACIPGSSSKQLFLFHVVOQPSGNEK 480
 Db 421 VVSFCMDHCKPVYIGTPVQIKLAOPPVLSOPACIPGSSSKQLFLFHVVOQPSGNEK 480
 QY 481 QVTTISHSSTLTQKCGOKTVPVTTIPTSQFPASITKOTITLPGNTLSIQASPTQKNR 540
 Db 481 QVTTISHSSTLTQKCGOKTVPVTTIPTSQFPASITKOTITLPGNTLSIQASPTQKNR 540
 QY 541 IKENWTSCFRDEDDINDVTSAGVNLNENACIATNSELVGLTIQSCDEPFIFALQ 600
 Db 541 IKENWTSCFRDEDDINDVTSAGVNLNENACIATNSELVGLTIQSCDEPFIFALQ 600
 QY 601 KRILIDKRRKHDITELNSDAVNLISQATQERLGLIEKLTALIAQHRMTYKASENYILCS 660
 Db 601 KRILIDKRRKHDITELNSDAVNLISQATQERLGLIEKLTALIAQHRMTYKASENYILCS 660
 QY 721 QHRDANTLATAAIGPRKRPLESGIEGLKDNLLASGTSSTLATKQLRRPRTTRICLDLI 780
 Db 721 QHRDANTLATAAIGPRKRPLESGIEGLKDNLLASGTSSTLATKQLRRPRTTRICLDLI 780
 QY 781 FCMQEREMKYSRALTALAK 801
 Db 781 FCMQEREMKYSRALTALAK 801

RESULT 2

T2D3_HUMAN STANDARD; PRT; 1083 AA.
 AC 000268; Q99721; Q9BX42; Q9BR40;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Transcription Initiation factor TFIIID 135 kDa subunit (TAFII-135)
 DE (TAFII135) (TAFII-130) (TAFII130).
 GN TAF4 OR TAF4A OR TAF2C1 OR TAF2C OR TAFII135 OR TAFII130.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97336072; PubMed=9192867;
 RA Mengus G., May M., Carre L., Chambon P., Davidson I.;

RT "Human TAF(II)135 potentiates transcriptional activation by the AF-2s
 of the retinoic acid, vitamin D3, and thyroid hormone receptors in
 mammalian cells.";
 RL Genes Dev. 11:1381-1395(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baggeley C.L.,
 Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
 Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
 Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 Coulson A., Coville G.J., Deaman R., Dhami P.D., Dunn M.,
 Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 Grahame D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 Levasseur M.H., Leverisa M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 Marsh V.L., Martin S.L., McConachie L.J., McElay K., McMurtry A.A.,
 Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.L.,
 Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,
 Stuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 Swan R.M., Symmore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
 Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
 Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 Rogers J.;
 RA "The DNA sequence and comparative analysis of human chromosome 20.";
 RT Nature 414:865-871(2001).
 RN [3]
 RP SEQUENCE OF 105-1083 FROM N.A. AND PARTIAL SEQUENCE.
 RX MEDLINE=97098442; PubMed=8942962;
 RA Tanese N., Saluja D., Vaasallo M.F., Chen J.-L., Admon A.;
 RT "Molecular cloning and analysis of two subunits of the human TFIIID
 complex: hTAFII130 and hTAFII100.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:13611-13616(1996).
 CC -1- FUNCTION: MAKES PART OF TFIIID IS A MULTIMERIC PROTEIN COMPLEX THAT
 PLAYS A CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS
 ACTIVATORS AND REPRESSORS. POTENTIATES TRANSCRIPTIONAL ACTION
 BY THE AF-2S OF THE RETINOIC ACID, VITAMIN D3 AND THYROID HORMONE.
 CC -1- SUBUNIT: TFIIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
 NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).
 CC -1- SUBCELLULAR LOCATION: Nuclear
 CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC EMBL: Y11354; CAA72189.1; -
 CC EMBL: AL137077; CAC36006.1; -
 CC EMBL: AL109911; CAC22312.2; -
 CC EMBL: U75308; AAC50901.1; -
 CC TRANSFAC: T02328; -
 CC GeneW: HGNC:11537; TAF4.
 CC MIM: 601796; -
 DR InterPro: IPR003894; TAF_hom.
 DR SMART: SM00549; TAFH.1.
 KW Transcription regulation; Nuclear protein.
 FT DOMAIN 39 42
 FT DOMAIN 52 57
 FT DOMAIN 98 101
 FT DOMAIN 142 148
 FT DOMAIN 268 275
 FT POLY-HIS.
 FT POLY-ALA.
 FT POLY-GLY.
 FT POLY-ALA.
 FT POLY-PRO.

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.